



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
-----------------	-------------	----------------------	---------------------	------------------

10/748,655

12/31/2003

Craig Nevill-Manning

0026-0049

2801

44989 7590 04/13/2009

HARRITY & HARRITY, LLP
11350 Random Hills Road
SUITE 600
FAIRFAX, VA 22030

EXAMINER

ALI, FARHAD

ART UNIT

PAPER NUMBER

2446

MAIL DATE

DELIVERY MODE

04/13/2009

PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No. 10/748,655	Applicant(s) NEVILL-MANNING ET AL.	
	Examiner FARHAD ALI	Art Unit 2446	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 30 December 2008.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-7, 9-12 and 14-22 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-7, 9-12, and 14-22 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date. _____ |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08) | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

Status of Claims:

Claims 1-7, 9-12, and 14-22 are pending in this Office Action.

Claims 1, 3-4, 6-7, 9, 12, 14, and 17 are amended.

Claims 21 and 22 are new.

Claim Rejections - 35 USC § 102

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

Claims 1-7, 9-12, and 14-22 are rejected under 35 U.S.C. 102(b) as being anticipated by Goodman (US 5,999,929 A).

Goodman teaches:

Claim 1

A method performed by a computer system, the method comprising:

receiving a first uniform resource locator (URL) (**Column 5 Lines 1-4, “the spider 14 uses URLs to identify Web pages to be retrieved for analysis”**);

selecting one or more parameters present in the first URL;

generating a plurality of different URLs having different parameter combinations of the one or more selected parameters;

retrieving using the first URL **(Column 5 Lines 5- “After the spider 14 receives a Web page for analysis, it caches the Web page locally within the link referral system”);**

retrieving content using the plurality of different URLs; comparing, by a processor of the computer system, the content retrieved using the first URL to the content retrieved using the plurality of different URLs; identifying, based on the comparing, one of the parameter combinations, that, when present in a particular URL, results in retrieving content that is approximately the same as the content corresponding to the first URL, the identifying being performed by the processor; and generating, by the processor, one or more URL rewrite rules based on the identified one of the parameter combinations **(Column 7-8 Lines 24-53, “In generating the URL re-write rules, the Web page analyzer 15 generally processes the URL from the outward most portions of the respective World Wide Web addresses, eliminating portions of the respective series, as defined by the separators, to determine candidate URLs. For the illustrative URL above, HTTP://www.netscape.com/ index.html”, candidate URLs will generally include, for example, eliminating portions from the beginning of the World Wide Web address”)**.

Claim 2

The method of claim 1, where the different parameter combinations include the first URL with no parameters, the first URL with each of the one or more parameters individually, and the first URL with combinations of the one or more parameters (**Column 7 Lines 24-28, “In generating the URL re-write rules, the Web page analyzer 15 generally processes the URL from the outward most portions of the respective World Wide Web addresses, eliminating portions of the respective series, as defined by the separators, to determine candidate URLs”, and also see Column 7 Lines 28-50).**

Claim 3

The method of claim 1, further comprising:

performing the receiving a first URL, the selecting one or more parameters present in the first URL, the generating a plurality of different URLs, retrieving content using the first URL, retrieving content using the plurality of URLs, the comparing the content, and identifying one of the parameter combinations, for multiple different first URLs that each include the one or more parameters; and

generating the one or more URL rewrite rules for the identified one of the parameter combinations for each of the first URLs (**See Claim 1 rejection**).

Claim 4

The method of claim 3, where the rewrite rules specify that parameters that do not occur in a threshold number of the identified one of the

Art Unit: 2446

parameter combinations are to be removed (**Column 8 Lines 30-33, “After generating the score, the Web page analyzer 15 will store the candidate re-write rule in the URL re-write rulebase 16B if the score is below a predetermined threshold value”).**

Claim 5

The method of claim 1, wherein each rewrite rule applies to a particular web site or web host (**Column 5 Lines 17-21, “To assist in the duplicate Web page consolidation operation, the Web page analyzer 15 develops the URL re-write rulebase 16B, which contains rules which are used by the Web page analyzer 15 to convert URLs to respective canonical forms”).**

Claim 6

The method of claim 1, where the identified one of the parameter combinations includes a minimum number of parameters with respect other ones of the parameter combinations (**Column 7 Lines 40-50, examples show removing portions from the “beginning” and “end” of the World Wide Web address without ever actually removing the first unique part of the URL).**

Claim 7

A method, performed by a computer system, for converting a uniform resource locator (URL) into a canonical form of the URL, the method comprising:

receiving a URL that refers to content and that includes a parameter set including at least one parameter **(Column 5 Lines 1-4, “the spider 14 uses URLs to identify Web pages to be retrieved for analysis”);**

selecting, by a processor of the computer system, a rewrite rule by receiving a plurality of URLs that include the parameter set and identifying parameters in the parameter set that do not result in retrieving substantially different content, when present in a URL **(Column 7-8 Lines 24-53, “In generating the URL re-write rules, the Web page analyzer 15 generally processes the URL from the outward most portions of the respective World Wide Web addresses, eliminating portions of the respective series, as defined by the separators, to determine candidate URLs. For the illustrative URL above, HTTP://www.netscape.com/ index.html”, candidate URLs will generally include, for example, eliminating portions from the beginning of the World Wide Web address”);**

applying, by the processor, the rewrite rule to the URL by removing the parameters that do not contribute to content from the URL; and outputting the rewritten URL as the canonical form of the URL **(Column 5 Lines 17-21, “To assist in the duplicate Web page consolidation operation, the Web page analyzer 15 develops the URL re-write rulebase 16B, which contains rules which are used by the Web page analyzer 15 to convert URLs to respective canonical forms”).**

Claim 9

The method of claim 7, where the identifying parameters in the parameter set includes; retrieving content corresponding to a sampled URL including a combination of parameters in the parameter set; and identifying the combination of parameters as corresponding to retrieved content, where the retrieved content is approximately the same as another retrieved content corresponding to another combination of parameters that includes a reduced number of parameters **(Column 8 Lines 1-9, “If the Web page analyzer 15 determines in step 2b that the URLs in the entry are not identical to each other, it (that is, the Web page analyzer 15) find the shortest substitution rule that textually rewrites the longer URL into the shorter URL. For example, the shortest rule to change http://www.netscape.com/index.html" to HTTP://netscape.com/index.html" is to replace "www." with "" (that is, delete "www."). This rule is now a "candidate" rewrite rule”)**.

Claim 10

The method of claim 9, where the combination of parameters includes at least one of the sampled URL with no parameters, the sampled URL with individual parameters, or the sampled URL with combinations of the at least one parameter **(Column 7 Lines 24-28, “In generating the URL re-write rules, the Web page analyzer 15 generally processes the URL from the outward most portions of the respective World Wide Web addresses, eliminating portions of the respective series, as defined by the separators, to determine candidate URLs”, and also see Column 7 Lines 28-50)**.

Claim 11

The method of claim 7, where the rewrite rule applies to a particular web site or web host **(Column 5 Lines 17-21, “To assist in the duplicate Web page consolidation operation, the Web page analyzer 15 develops the URL re-write rulebase 16B, which contains rules which are used by the Web page analyzer 15 to convert URLs to respective canonical forms”)**.

Claim 12

One or more devices comprising:

at least one fetch bot to download content on a network from locations specified by uniform resource locators (URLs) **(Column 4 Lines 60-65, “spider”)**;

a content manager configured to extract URLs from the downloaded content **(Column 5 Lines 5-10, “Web page analyzer”)**;

a rewrite component to receive a URL that refers to content and that includes a parameter set including at least one parameter, apply a predetermined rewrite rule to the URL that removes the at least one parameter from the URL when the at least one parameter does not affect the content referred to by the URL, where the predetermined rewrite rule is determined by receiving a plurality of URLs that include the parameter set and identifying parameters in the parameter set that do not result in retrieving substantially different content, when present in a URL; and output the rewritten URL as the canonical form of the URL **(Column 5 Lines 17-21, “To assist in the duplicate**

Web page consolidation operation, the Web page analyzer 15 develops the URL re-write rulebase 16B, which contains rules which are used by the Web page analyzer 15 to convert URLs to respective canonical forms”); and a URL manager configured to store the canonical form of the URL (Column 5 Lines 30-33, “The Web page analyzer 15 stores information regarding the identifications for the various classes and the Web page assignment information in the link class database 17”).

Claim 14

The one or more devices of claim 12, where the identifying parameters in the parameter set includes; retrieving content corresponding to a sampled URL including a combination of parameters in the parameter set; and identifying the combination of parameters as corresponding to retrieved content, where the retrieved content is approximately the same as another retrieved content corresponding to another combination of parameters that includes a reduced number of parameters **(Column 8 Lines 1-9, “If the Web page analyzer 15 determines in step 2b that the URLs in the entry are not identical to each other, it (that is, the Web page analyzer 15) find the shortest substitution rule that textually rewrites the longer URL into the shorter URL. For example, the shortest rule to change http://www.netscape.com/index.html” to “HTTP://netscape.com/index.html” is to replace “www.” with “” (that is, delete “www.”). This rule is now a “candidate” rewrite rule”).**

Claim 15

The one or more devices of claim 14, where the combination of parameters includes at least one of the sampled URL with no parameters, the sampled URL with individual parameters, or the sampled URL with combinations of the at least one parameter (**Column 7 Lines 24-28, “In generating the URL re-write rules, the Web page analyzer 15 generally processes the URL from the outward most portions of the respective World Wide Web addresses, eliminating portions of the respective series, as defined by the separators, to determine candidate URLs”, and also see Column 7 Lines 28-50).**

Claim 16

The one or more devices of claim 12, where each rewrite rule applies to a particular web site or web host (**Column 5 Lines 17-21, “To assist in the duplicate Web page consolidation operation, the Web page analyzer 15 develops the URL re-write rulebase 16B, which contains rules which are used by the Web page analyzer 15 to convert URLs to respective canonical forms”).**

Claim 17

A system comprising:

means for receiving a first uniform resource locator (URL) including one or more parameters (**Column 5 Lines 1-4, “the spider 14 uses URLs to identify Web pages to be retrieved for analysis”);**

means for retrieving content corresponding to the first URL (**Column 5 Lines 5-“After the spider 14 receives a Web page for analysis, it caches the Web page locally within the link referral system”**);

means for retrieving content corresponding to a plurality of URLs having different parameter combinations of the one or more parameters; means for identifying the parameter combination from the plurality of URLs that corresponds to content that is approximately the same as the content corresponding to the first URL and that contains a minimum number of parameters compared to other parameter combinations; (**Column 7-8 Lines 24-53, “In generating the URL re-write rules, the Web page analyzer 15 generally processes the URL from the outward most portions of the respective World Wide Web addresses, eliminating portions of the respective series, as defined by the separators, to determine candidate URLs. For the illustrative URL above, HTTP://www.netscape.com/ index.html”, candidate URLs will generally include, for example, eliminating portions from the beginning of the World Wide Web address”**); and

means for generating one or more URL rewrite rules based on the identified parameter combination (**Column 5 Lines 17-21, “To assist in the duplicate Web page consolidation operation, the Web page analyzer 15 develops the URL re-write rulebase 16B, which contains rules which are used by the Web page analyzer 15 to convert URLs to respective canonical forms”**).

Art Unit: 2446

A computer-readable memory device including programming instructions executed by a processor, the programming instructions comprising:

instructions for receiving a first uniform resource locator (URL) including one or more parameters **(Column 5 Lines 1-4, “the spider 14 uses URLs to identify Web pages to be retrieved for analysis”);**

instructions for retrieving content corresponding to the first URL **(Column 5 Lines 5- “After the spider 14 receives a Web page for analysis, it caches the Web page locally within the link referral system”);**

instructions for retrieving content corresponding to a plurality of URLs having different parameter combinations of the one or more parameters; instructions for identifying the parameter combination from the plurality of URLs that corresponds to content that is approximately the same as the content corresponding to the first URL and that includes a minimum number of parameters **(Column 7-8 Lines 24-53, “In generating the URL re-write rules, the Web page analyzer 15 generally processes the URL from the outward most portions of the respective World Wide Web addresses, eliminating portions of the respective series, as defined by the separators, to determine candidate URLs. For the illustrative URL above, HTTP://www.netscape.com/ index.html”, candidate URLs will generally include, for example, eliminating portions from the beginning of the World Wide Web address”); and**

instructions for generating one or more URL rewrite rules based on the identified parameter combination (**Column 5 Lines 17-21, “To assist in the duplicate Web page consolidation operation, the Web page analyzer 15 develops the URL re-write rulebase 16B, which contains rules which are used by the Web page analyzer 15 to convert URLs to respective canonical forms”**).

Claim 19

The system of claim 17, where the parameter combination comprises one of the first URL with no parameters, the first URL with each of the one or more parameters individually, or the first URL with combinations of the one or more parameters (**Column 7 Lines 24-28, “In generating the URL re-write rules, the Web page analyzer 15 generally processes the URL from the outward most portions of the respective World Wide Web addresses, eliminating portions of the respective series, as defined by the separators, to determine candidate URLs”, and also see Column 7 Lines 28-50**).

Claim 20

The computer-readable memory device of claim 18, where the instructions for receiving a first URL, the instructions for retrieving content corresponding to the first URL, the instructions for retrieving content corresponding to a plurality of URLs, and the instructions for identifying the parameter combination are performed for multiple first URLs, each first URL including the one or more parameters (**See claim 18 rejection**), and where the one or more URL rewrite rules specify that parameters that do not occur

in a threshold number of the identified parameter combinations are to be removed
(Column 8 Lines 30-33, “After generating the score, the Web page analyzer 15 will store the candidate re-write rule in the URL re-write rulebase 16B if the score is below a predetermined threshold value”).

Claim 21

The system of claim 17, further comprising: means for determining whether the content that corresponds to the plurality of URLs is approximately the same as the content that corresponds to the first URL using a similarity hash function **(Hash function is a well known function for comparing documents. Applicant admits in paragraph [0041] of specification “A document having “approximately the same content” as another document may be determined using any of a number of known document comparison techniques, such as comparison techniques based on a similarity hash”).**

Claim 22

The computer-readable memory device of claim 18, where the rewrite rules specify that parameters that do not occur in a threshold number of the identified parameter combinations are to be removed **(Column 8 Lines 30-33, “After generating the score, the Web page analyzer 15 will store the candidate re-write rule in the URL re-write rulebase 16B if the score is below a predetermined threshold value”).**

Response to Arguments

2. Applicant's arguments filed 12/30/2008 have been fully considered but they are not persuasive.

In regards to utilizing the Goodman reference, the examiner has considered the pending claims and upon further consideration found the Goodman reference to be of particular relevance. The Applicant's arguments are addressed below.

Applicant has argued that Goodman does not disclose or suggest retrieving content using a plurality of different URLs having different parameter combinations of one or more selected parameters, as recited in amended claim 1.

The Examiner respectfully disagrees. In Column 7 lines 24-32, Goodman teaches "In generating the URL re-write rules, the Web page analyzer 15 generally processes the URL from the outward most portions of the respective World Wide Web addresses, eliminating portions of the respective series, as defined by the separators, to determine candidate URLs". The examiner asserts that Goodman is receiving content, wherein eliminating portions of the respective series to determine candidate URLs is analogous to the applicants retrieving content using a plurality of different URLs having different parameter combinations of one or more selected parameters as claimed.

Applicant has argued that Goodman does not disclose or suggest identifying, based on comparing content retrieved using a first URL to content retrieved using a

Art Unit: 2446

plurality of different URLs, a parameter combination, that, when present in a particular URL, results in retrieving content that is approximately the same as content corresponding to the first URL, the identifying being performed by a processor, as also recited in amended claim 1.

Applicants arguments regarding Independent claim 7 are similarly addressed below.

The Examiner respectfully disagrees. Goodman teaches in Column 5 Lines 17-21, "To assist in the duplicate Web page consolidation operation, the Web page analyzer 15 develops the URL re-write rulebase 16B, which contains rules which are used by the Web page analyzer 15 to convert URLs to respective canonical forms" and in Column 8 Lines 1-9, "If the Web page analyzer 15 determines in step 2b that the URLs in the entry are not identical to each other, it (that is, the Web page analyzer 15) find the shortest substitution rule that textually rewrites the longer URL into the shorter URL. For example, the shortest rule to change `http://www.netscape.com/index.html` to `"HTTP://netscape.com/index.html"` is to replace `"www."` with `""` (that is, delete `"www."`). This rule is now a "candidate" rewrite rule". The examiner asserts that this operation is analogous to the applicants identifying a parameter combination that results in retrieving content that is approximately the same as content corresponding to the first URL.

Conclusion

3. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

Art Unit: 2446

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to FARHAD ALI whose telephone number is (571)270-1920. The examiner can normally be reached on Monday thru Friday, 7:30am to 5:00pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Jeffrey C. Pwu can be reached on (571) 272-6798. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Art Unit: 2446

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Farhad Ali/
Examiner, Art Unit 2446

/Jeffrey Pwu/
Supervisory Patent Examiner, Art Unit 2446